DOCUMENT RESUME

ED 243 968 TM 840 295

AUTHOR Hoge, Robert D.
TITLE Observational Measures of Classroom Behavior: A Critical Examination.
PUB DATE [82]
NOTE 49p.
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Behavior Modification; *Classroom Observation Techniques; Elementary Secondary Education; *Evaluation Methods; Intervention; Outcomes of Education; Research Design; *Research Methodology; Research Problems; School Surveys; Student Behavior; *Validity

ABSTRACT

The focus of this review is on observational measures of pupil classroom behaviors. Two issues are raised in connection with these measures. First, there is a survey of the types of observation schedules employed in recent classroom intervention research. Second, there is an evaluation of the validity of these behavioral measures. That evaluation is based on an examination of empirical data, and the data are drawn from three types of analyses: cases where (1) observational measures were related to alternative measures of the behaviors, (2) observational measures were related to performance indices within correlational designs, and (3) observational measures were related to performance measures in experimental designs. The outcomes of the survey and the evaluation are used to derive some recommendations relevant to the use of these measures in applied and research settings and some recommendations regarding directions for future research with the measures. (Author)

Observational Measures of Classroom Behavior: A Critical Examination

Robert D. Hoge

Carleton University

Running Head: OBSERVATIONAL MEASURES OF CLASSROOM BEHAVIOR

Abstract

The focus of the review is on observational measures of pupil classroom behaviors. Two issues are raised in connection with these measures. First, there is a survey of the types of observation schedules employed in recent classroom intervention research. Second, there is an evaluation of the validity of these behavioral measures. That evaluation is based on an examination of empirical data, and the data are drawn from three types of analyses: cases where (a) observational measures were related to alternative measures of the behaviors, (b) observational measures were related to performance indices within correlational designs, and (c) observational measures were related to performance measures in experimental designs. The outcomes of the survey and the evaluation are used to derive some recommendations relevant to the use of these measures in applied and research settings and some recommendations regarding directions for future research with the measures.

Observational Measures of Classroom Behavior: A Critical Examination

The focus of this paper is on observational measures of pupil classroom behavior as employed within behavior modification research. Such measures have been the object of some recent theoretical and methodological attention. For example, Wasik and Loven (1980) have presented an examination of reliability problems associated with the measures and Hoge and Luce (1979) have presented a summary of the achievement correlates of the measures.

What has been missing, however, is a broad-based survey and evaluation of these measures. This review is designed to provide such an examination, and the issue is approached from two directions. First, there is a description of the types of observational measures employed in recent behavior modification research. Second, there is an evaluation of the measures. This evaluation focuses on questions about the validity of the measures and is based on an examination of empirical data. The issues raised in this evaluation are shown to relate to some key assumptions which are made about the measures as they are used in applied and research settings.

Description of the Measures

The description of the measures is based on a survey designed to uncover the various types of schedules employed in recent behavioral intervention studies. The survey itself is based on a review of studies published between 1977 and 1983 which involved a classroom intervention procedure and which included an observational measure of pupil classroom behavior as a criterion or dependent measure.[1] The purpose of this

survey is to familiarize researchers and practitioners with the range of category systems being employed in this research and with certain relevant features of those systems.

Table 1 contains a summary of the various category systems being used in these studies. It can be seen from the table that a wide variety

---

Insert Table 1 about here

---

of observation schedules have been developed. There are, however, two bases for characterizing these systems which have some relevance for their use in applied and research settings. These bases relate, as is shown in Figure 1, to the breadth and specificity of the systems.

Insert Figure 1 about here

The breadth dimension is described at one extreme by those schedules which provide for a focus on a limited range of behaviors. An example is that employed by Jones, Fremouw, and Carples (1977) with its two categories of "talk to neighbor" and "out of seat". At the other extreme are the schedules which include a broader range of classroom behaviors. An example is that employed by Hops and others (1978) with its 13-categories of behavior. The decision to employ a narrow or broad schedule will depend largely on the purposes of the assessment in a particular situation. There are, however, some practical considerations associated with the decision. In general, increases in the breadth of a system are accompanied by increased problems with observer training, observer agreement, etc. (Rosenshine & Furst, 1973).

The second dimension identified in Figure 1 relates to the specificity of categories represented within the system. At one extreme here are the global categories as represented, for example, in those labeled "on task" or "appropriate behavior". At the other extreme are the more specific molecular categories such as "out of chair" or "look around".

This distinction between specific and global observation categories has some important implications. First, there are implications for the level of inference represented in the category. As indicated in Figure 1, levels of inference are typically higher with the global categories than with the specific categories. Thus, a higher level of observer judgment is called for in the case of the category "inappropriate classroom behavior" than with a category such as "out of seat". The level of inference represented in the measure is important because of its implications for the use of the system — training and application are usually easier with the low inference measures — and for questions of reliability and validity (see Cone, 1982; Dunkin & Biddle, 1974; Rosenshine & Furst, 1973).

Another implication associated with the specific vs. global distinction concerns the precision of operational definitions. The provision of precise operational definitions of response categories is essential for all types of measures. The need is particularly acute for the global measures which involve high levels of inference on the observer's part and for which there is considerable latitude of interpretation. This leads to an important observation respecting the various systems described in Table 1. While the operational definitions

associated with the specific category systems tend to be complete and precise, there is often a lack of precision and consistency associated with the global systems. For example, Cameron and Robinson (1980) use the global category "on task" behavior, and they define that category as "... appropriate engagement in assigned tasks, including working individually with the teacher, waiting with hand raised, organizing materials at start of lesson, use of eraser to correct answers, checking answers, and recording results" (p. 408). Although this definition identifies some behaviors likely associated with "on task" behaviors, there would remain considerable latitude in connection with the decision to categorize a behavior as "on task" or "off task". These considerations probably do not affect the use of these measures within individual studies. They are, however, of some relevance when it comes to generalizing across and beyond studies. Unless categories are defined precisely and consistently, there is simply no basis for generalization (Cone & Foster, 1982; Dunkin & Biddle, 1974; Hartmann, Roper, & Bradford, 1979; Karweit & Slavin, 1982; Klein, 1979).

Evaluation of the Measures

The preceding section has provided a survey of the available measures and some evaluative comments respecting their format and content. The concern in this section is with the measurement properties of these observational measures. The traditional psychometric model specifies two bases for evaluating psychological measures, in terms of reliability and validity. The issue of the reliability of these measures of pupil classroom behaviors has been dealt with in a recent review of Waskin and Loven (1980) and will not be discussed further here.

The issue of the validity of the measures has, on the other hand, been somewhat neglected, as is often the case with behavioral measures. Questions of validity are, however, important. The use of these observational measures in applied and research settings is based on certain assumptions regarding their meaning and relevance, and it is important to know to what extent these assumptions are being met (Cone, 1982; Emery & Marholin, 1977; Foster & Cone, 1980; Herbert & Attridge, 1975; Rosenshine & Furst, 1973).

Validity Paradigms

Just as there is some controversy over the appropriateness of the psychometric model for behaviorist methodology (e.g., Hartmann et al., 1979; Nelson, 1983), so there has also been some ambiguity associated with the way in which the model has been applied to the assessment of behavioral measures. Cone (1982), however, has recently presented a useful system for applying the validity construct to behavioral measures, and his system will be used to organize the present discussion.

Cone (1982) includes four forms of validity within his system. The first type is content validity, and this refers to the extent to which components of the observational measure correspond in a logical way to the behaviors or theoretical constructs presumably being measured by the instrument. For example, do the behavior categories making up the measure of "inappropriate classroom behavior" truly reflect that behavioral domain? The second form of validity specified in the system is criterion-related validity, and this form is represented where relations are established between the observational measure and some alternative measure. An example would be the case where an observational measure of

"on task" behavior is correlated with an index of academic achievement. Construct validity, the third form represented in the system, refers to the extent to which scores from the observational measure correspond to theoretically relevant measures. As Cone notes, this form of validity is applicable where one is concerned with establishing the meaning of deductively formed constructs. Thus, efforts to relate a composite measure of "deviant" behavior to alternative indices of deviant behaviors would correspond to construct validity. The fourth form of validity represented in the system is termed treatment validity and refers to the extent to which the use of the measure is associated with intervention outcomes.

Cone (1982) has also specified in his system two dimensions which are relevant to the interpretation of validity data. The first of these relates to the subject matter of the observational assessment. The basic distinction underlying this dimension relates to a focus on discrete and observable behaviors versus a focus on superordinate constructs derived from the discrete measures and which usually relate to psychological traits. The second dimension relates to the purposes of the assessment, with the basic distinction here between applied-practical uses of observational data and scientific-theoretical uses of the data. The relevance of these distinctions for the consideration of validity will be shown below.

The assessment of the first form of validity specified in the system, content validity, is largely dependent on intuitive and deductive processes. The assessment of the three other forms of validity, on the other hand, depends upon empirical procedures. The purpose of this

section of the paper is to consider the empirical data which are available with respect to the pupil behavior measures. These data derive from three types of studies: cases where (a) the classroom observation measure is related to teacher rating measures of the same or alternative behaviors, (b) the observational measures are related to measures of academic performance within correlational designs, and (c) the measures are related to academic performance within experimental designs. The relevance of these data for the validity of the observational measures is then considered in terms of the system developed by Cone (1982).

The Validity Data

Relations with teacher judgment measures. The studies summarized here have all reported data on relations between observational measures of pupil classroom behavior and alternative measures derived from teacher ratings. This type of analysis bears most directly on the issue of criterion-related validity. The information is of particular relevance where measures are used within applied-practical contexts because, in those contexts, links are often assumed between the observational measure and criterion measures. It is in this type of context that, for example, a demonstration of significant relations between an observational measure of "on task" behavior and a teacher rating measure of classroom adjustment would be of interest.

Data on relations between observational measures and teacher judgment measures may also, under some circumstances, be relevant to construct validity. This would be the case where the observational measure is used as an index of a hypothetical construct, whether a psychological trait or some other type of construct, and the criterion measure represents an

alternative index of that construct. For example, a demonstration of convergence between a composite observational measure of hyperactivity and teacher ratings of hyperactivity would reflect on the construct validity of the observational measure (Cone, 1979, 1982; Gresham, 1982; Messick, 1981).

Two of the more direct efforts to assess the validity of an observation measure may be seen in studies reported by Hudgins (1967) and Blunden, Spring, and Greenberg (1974). The Hudgins study involved relating an observational measure of pupil attentiveness to teacher ratings of attentiveness. Separate correlations were reported for each of nine teaches, and, while there was some variability among the teachers, the correlations were generally strong and statistically significant (median $\underline{r}$ = .65). Blunden et al. (1974) collected observational data in terms of 10 categories of classroom behavior, and they related those measures to teacher ratings on the 10 behavioral dimensions. They reported generally nonsignificant relations between the corresponding measures.

Green, Beck, Forehand, and Vosk (1980) and Lahey, Green, and Forehand (1980) employed a behavioral observation schedule first developed by Hartup, Glazer, and Charlesworth (1967). The schedule involved the following behavioral categories: (a) "alone and on task", (b) "interacting with teacher" (positive or negative), (c) "interacting with peer" (positive or negative). Data from both studies revealed only weak relations between these observational categories and clinical groupings of subjects formed on the basis of teacher rating, using similar categories.

Studies reported by Bolstad and Johnson (1977), Nelson (1971), Werry

and Quay (1969), and Zentall (1980) employed designs similar to those used in the two studies just described, but these researchers obtained somewhat more positive results for the behavioral measures. Werry and Quay (1969), for example, contrasted groups of conduct problem and normal children in terms of seven categories of deviant behaviors and three categories of attentive behaviors. Significant differences were obtained between the teacher designated normal and control groups for most of the behavioral categories. Similar positive results were reported by Bolstad and Johnson (1977), Nelson (1971), and Zentall (1980).

The 20-item behavioral schedule employed by Whalen et al. (1979) was described in Table 1. Those researchers provided some information with respect to the validity of their schedule by reporting correlations between category scores and a total score derived from the Conners Abbreviated Symptom Questionnaire. The latter is a teacher judgment measure of the hyperactive syndrome. Separate correlations were reported for each of the 21 behavioral categories; 11 of those 21 correlations were statistically significant, and the range of correlations was from .25 to .78. As might be expected, the strongest correlations were between the hyperactivity score and the behavioral dimensions of "task attention", "noise", "disruption", and "inappropriate stand out".

One of the most recent and most interesting developments in this area is represented in the work of Abikoff, Gittelman-Klein, and Klein (1977, 1980). These researchers are in the process of developing an observation schedule appropriate for the identification and assessment of hyperactive children. The most recent version of this observation schedule contains 14 categories relating to specific aspects of classroom behavior (e.g.,

"off task", "noncompliance", "verbal aggression to teacher"). Much of the work with this schedule has been directed toward reliability assessments, but some information has been presented relevant to the validity of the schedule. These analyses concerned the ability of the behavioral categories to discriminate between groups of hyperactive and normal children, with the latter grouping based on teacher and parent ratings. Data from the two studies indicated that most of the specific behavioral categories were capable of discriminating between the two groups of subjects. The researchers have also begun to explore the formation of composite behavioral categories. By way of illustration, they have shown that the combination of the two categories "interference" and "off task" produced an 80% accuracy rate in the prediction of category membership. While some questions have been raised about the reliability and validity procedures employed in these studies (Cone, 1982; Haynes & Kerns, 1979), this work does indicate the type of careful instrument development needed in this area.

The set of studies reviewed in this section yielded somewhat mixed results. There were positive findings here; that is, there were successful efforts to relate a behavioral observation measure to alternative measures (Hudgins, 1967; Whalen et al., 1979) or to show that the behavioral measure could discriminate among clinical groupings of subjects (Abikoff et al., 1977, 1980; Bolstad & Johnson, 1977; Nelson, 1971; Werry & Quay, 1969; Zentall, 1980). These results relate clearly to the criterion-related validity of the observational measures, and they are such as to increase our confidence that we are dealing here with meaningful and relevant measures. Further, the results of the Abikoff et

al. (1977, 1980) and Whalen et al. (1979) studies have some bearing on the construct validity to the extent that they showed significant relations between alternative measures of similar hypothetical constructs.

There were, on the other hand, some negative results here as well. It may be noted, first, that even in those cases where significant results were reported, the magnitudes of relations tended to be rather low. Second, there were some clear cases of failures to establish relations between the observational and judgmental measures (Blunden et al., 1974; Green et al., 1980; Lahey et al., 1980). There is a third point to be made here as well. Those cases where positive results were reported involved, with the exception of the Hudgins (1967) study, relating specific observational categories to global criterion measures. There were no cases reported where specific observational categories were related to parallel specific judgmental categories. This is an important point because there are many cases, involving both applied and research contexts, where the validity of specific behavioral categories is assumed (Cone, 1981, 1982; Cone & Foster, 1982).

There are also some cautions which should be introduced with respect to the interpretation of the negative results. Such negative results may, in fact, reflect a lack of validity in the observational measures. There are, however, alternative interpretations. Thus, these failures may reflect inadequacies in the judgmental measures. Hoge (in press) has recently shown that some limitations exist with respect to the reliability and validity of teacher judgment measures. A second alternative is that the negative results may simply reflect a basic lack of correspondence between observational and judgmental types of measures. This possibility

has been discussed by a number of writers, including Cairns and Green (1979) and Cone and Foster (1982). The existence of these alternative interpretations does not mean, of course, that we can ignore the discrepant results. It does mean, though, that they should be interpreted with some caution.

Relations with achievement measures: correlational designs. The studies reviewed in this section all included analyses in which observational measures of pupil classroom behavior were related to indices of academic achievement within correlational designs. Data from these analyses may be viewed as bearing directly on the issue of criterion-related validity. This type of validity information is especially relevant within certain applied-practical uses of the measures: there is often an assumption made there that links exist between the classroom behaviors being assessed by the measures and academic achievement (Hoge & Luce, 1979; Lipe & Jung, 1971; Nelson & Hayes, 1979; Sherman & Bushell, 1975). In fact, the well-known debate between Winett and Winkler (1972) and O'Leary (1972) revolved to a large extent around the academic relevance of the pupil behaviors being selected for assessment and modification. This is not to say that links are always assumed between these behaviors and academic achievement or that enhancing achievement constitutes the only basis for selecting behaviors for modification. Still, there are many cases in which the links are assumed to exist, and it is for this reason that this type of validity is so important.

Lahaderne (1968) reported one of the earliest studies on this issue. She employed an observation schedule based on two broad categories of classroom behavior, "attentive" and "inattentive". This measure is

conceptually similar to the "on task" measure used in many studies included in the survey. Lahaderne reported correlations between standardized achievement tests scores and observation scores across male and female pupils and across a variety of achievement areas. All correlations were statistically significant, and the median correlation was $\underline{r} = .47$ for the "attentive" category and $\underline{r} = .45$ for the "inattentive" category. Luce and Hoge (1978) and Samuels and Turnure (1974) employed the same observational schedule, and they too reported significant relations between the attentiveness measure and achievement indices. However, a similar type of measure was employed by Hall, Huppert, and Levi (1977), and they failed to obtain significant correlations between the behavioral and achievement indices.

Another group of studies employed observation schedules which provided for a focus on a larger number of specific classroom behaviors (Cobb, 1970, 1972; Soli & Devine, 1976). The schedules used in these studies varied somewhat in the number and labeling of categories, but they all derived from the survival skill measure first reported by Cobb (1969). Variations of this schedule were seen in the Greenwood et al. (1977a, 1977b) and Hops et al. (1978) studies which were described in Table 1.

All three researchers reported significant correlations between specific behavior categories and achievement indices. However, a close examination of their results reveal three points. First, while the correlations were often statistically significant, their magnitudes were generally low. Second, there were usually as many nonsignificant correlations as significant ones. Third, efforts to cross-validate the correlations generally met with limited success. These points can be

illustrated with data from the Cobb (1972) study. Thirty-two correlations between behavior categories and achievement indices were reported (eight behavior categories x two achievement areas x two schools). Fifteen of those correlations were statistically significant, and the median of all correlations was $r = .25$. Further, there were some rather striking discrepancies in patterns of relations across the two schools. For example, the "out-of-chair" category showed a significant positive correlation with arithmetic achievement in the case of one school and a significant negative relation with arithmetic achievement in the case of the second school. Similar kinds of results were reported in the other two studies.

While the efforts to correlate individual category scores with achievement indices in these three studies did not yield very strong results, the outcomes of multiple regression analyses, involving the formation of composite survival skills, yielded higher levels of predictability. For example, Cobb (1970) reported an R of .70 for the prediction of reading achievement from behavioral data in the case of one school. The most heavily weighted categories in that equation were "talk to peer positive", "compliance", and "approval". A second example may be found in the Soli and Devine (1976) study where an R of .45 was reported for the prediction of mathematics achievement with the following response categories most heavily weighted in the equation: "interaction with peer positive", "not attending", "self stimulation", and "attending".

A final study to be mentioned in this section employed a somewhat different approach to the issue. McKinney, Mason, Perkerson, and Clifford (1975) collected behavioral observations in terms of a 27-item observation

schedule, with the items on that schedule providing for a focus on relatively specific aspects of classroom behavior. Data collected with the schedule were factor analyzed, and the analysis yielded a set of 12 factors. These factor scores were then used in multiple regression equations as predictors of standardized achievement test scores. Significant levels of prediction were obtained for three separate multiple regression analyses. By way of illustration, an R of .63 was obtained for the prediction of achievement, with the factors labeled Distractible Behavior, Passive Responding, and Dependency showing the heaviest weightings in the equations.

The set of analyses reviewed in this section provided information on relations between behavioral observation measures and performance indices. The results presented a rather mixed picture. With the exception of the Hall et al. (1977) study, significant relations were reported between global attention measures and achievement indices (Lahaderne, 1968; Luce & Hoge, 1978; Samuels & Turnure, 1974).[2] This is an encouraging finding since that type of measure corresponds to the "on task" measure so widely used in behavior modification research. Weaker support for criterion-related validity was generally found with the specific behavior categories. However, when these specific categories were combined into composites through statistical means, as was the case with the final four studies reviewed, higher levels of prediction were shown. These efforts at developing composite indices through multiple regression or factor analytic procedures were too few in number to reach any firm conclusion about optimal combinations of specific behaviors, but this does indicate a

promising direction for future research on the formation of composites and the identification of critical academic survival skills.

There is another issue raised in these studies which bears mention, and this concerns the existence of moderator variables. There has been some evidence that the classroom behavior-academic achievement relation may vary as a function of contextual or subject variables (Hoge & Luce, 1979). For example, Cobb (1970) found a higher correlation between classroom behavior and achievement in the case of boys than in the case of girls; and, further, he found somewhat different behavioral indices entering regression equations in the two cases. To take another example, Soli and Devine (1976) found different behaviors predictive of arithmetic and reading achievement. The findings here are too few in number to warrant any firm conclusions about which variables may function as moderators. The approach is, however, a useful one since it may lead to the identification of critical contextual and subject variables, and, that, in turn, would have important implications for the selection of behaviors for modification.

Relations With Achievement Measures (Experimental Designs)

The correlational studies just reviewed are of interest because they provide us with information about the extent to which links exist between classroom behaviors and academic performance. However, the assumption often made in using the measures is that the classroom behaviors are related in a causal fashion to academic achievement, and the correlational studies are not capable of providing information on that point. The assumption of causality can be addressed only through experimental studies, and the relevant experimental studies are reviewed in this section.

All of the investigations reviewed here provided information on the behavior-achievement link within the context of an experimental design. One set of studies included those in which there was an effort to manipulate classroom behaviors directly, with indices of classroom behavior and indices of achievement serving as dependent variables. This type of design provides direct information regarding the extent to which a functional or a causal link exists between classroom behaviors and achievement. To the extent that the manipulation of classroom behavior leads to alterations in achievement, we may say that evidence for such links exists. A second set of studies includes those in which the experimental manipulation was directed toward academic performance, with indices of classroom behavior and performance serving as dependent measures. Results from these studies would seem to bear somewhat less directly on the assumption of a causal link between classroom behaviors and achievement, but the results are informative so far as the issue is concerned, and the studies are included here. A third category of study, that in which both academic behaviors and performance are manipulated, is also included in the discussion.[3]

It should be noted at the outset that two different types of criterion measure are represented in this research. Some researchers have used standardized achievement tests as the criterion measure, while in other cases criterion-referenced indices (e.g., number of problems attempted, percentage of correct answers) were employed. These indices tap somewhat different aspects of performance as Greenwood et al. (1979) have pointed out. The two types of indices also involve different timings

of data collection, with the criterion-referenced measures usually collected concurrent with the manipulation and the standardized measures collected prior to and following the manipulation.

Four studies have been reported in which the experimental manipulation was directed toward changes in academically relevant classroom behaviors. The effects of the manipulation were assessed against standardized achievement test scores in three of the studies (Cobb & Hops, 1973; Greenwood et al., 1977a; Greenwood et al., 1979) and against criterion referenced performance measures in the case of the fourth study (Friedling & O'Leary, 1979).

The Cobb and Hops (1973), Greenwood et al. (1977a), and Greenwood et al. (1979) studies all involved experimental conditions in which teachers in regular classrooms attempted to increase levels of appropriate classroom behaviors (e.g., "attending", "compliance") through systematic reinforcement. The manipulations produced significant effects on classroom behavior in all three studies; in other words, levels of appropriate behaviors did increase as a function of the selective reinforcement programs. However, the effects on the achievement measure were mixed. While Cobb and Hops (1973) were able to show significantly greater achievement gains for their experimental group relative to a control group, neither Greenwood et al. (1977a) nor Greenwood et al. (1979) were able to demonstrate very strong effects of the behavioral intervention on achievement. The fourth study in this category, Friedling and O'Leary (1979), also involved the manipulation of classroom behaviors (within an experimental classroom setting in this case), but these researchers employed indices of quantity of problems attempted and percentages of

correct solutions as performance measures. Here, too, essentially negative results were obtained, for, while the behavioral intervention produced significant behavioral change, there were no significant effects for the performance measure.

Another category of study involves the case where there was an effort to modify both classroom behaviors and academic performance within separate experimental conditions. Two of the six studies in this category employed standardized achievement tests as performance measures (Hops & Cobb, 1974; Walker & Hops, 1976). In both cases the focus of the behavioral intervention was on the enhancement of appropriate classroom behaviors. The alternative intervention effort was directed toward the development of basic reading skills in the case of the Hops and Cobb (1974) study and toward specific aspects of performance (e.g., problems completed) in the case of the Walker and Hops (1976) investigation. It was shown that both types of intervention were effective in producing both significant behavior change and significant achievement change relative to control groups receiving no interventions.

The remaining four studies in this category also contrasted conditions in which the focus of intervention was on behavior change with conditions in which the focus of intervention was on academic performance (Ferritor, Buckholdt, Hamblin, & Smith, 1972; Hay et al., 1977; Hundert, Bucher, & Henderson, 1976; Marholin & Steinman, 1977). These studies, too, included observational measures of classroom behavior as dependent variables. They differ from the two previous studies in that the performance measures in these cases were based on criterion-referenced indices rather than standardized achievement tests. The behavioral interventions in these

studies produced significant effects on classroom behavior. For example, Hay et al. (1977) were able to show significant increases in levels of "on task" behavior within the modification condition. However, in all of these cases nonsignificant effects were obtained for performance measures. These researchers, in other words, were able to show that a behavioral intervention will produce significant changes in survival skills but will have no impact on academic performance. It may also be noted here that in three of these studies, Hay et al. (1977), Hundert et al. (1976), and Marholin and Steinman (1977), the academic performance manipulation produced significant effects for both the behavioral indices and the performance studies.[4]

It was argued earlier that this latter type of finding, involving a demonstration of a significant effect of a performance manipulation on behavior, relates only indirectly to the issue of a causal link between classroom behavior and academic performance. The results are noted here for the sake of completeness, and for the same reason it may be noted that another set of studies exists in this literature, studies in which efforts were made to modify only academic performance but including measures of performance and behavior (Ayllon & Roberts, 1974; Ayllon, Layman, & Kundel, 1975; Broughton & Lahey, 1978; Center et al., 1982; Kirby & Shields, 1972; Winett & Roach, 1973). All of these investigators were able to show that the academic performance intervention was effective in producing changes in both performance and behavior.

It was argued earlier that the use of these behavioral measures in applied and research contexts is sometimes based on the assumption of causal links between the behaviors and academic achievement. The evidence

on that assumption is mixed. It can be observed, first, that none of the studies employing a criterion referenced performance measure was able to show a link between behaviors and performance. Most of those researchers were able to demonstrate that a behavioral intervention will produce the desired behavior changes, but there were no corresponding changes in performance. The studies employing standardized achievement tests as dependent measures yielded some positive results, but here too the outcomes were mixed. Thus, while Cobb and Hops (1973), Hops and Cobb (1974), and Walker and Hops (1976) showed relatively strong effects on achievement for a behavioral intervention, other researchers employing this design obtained rather weaker effects.

The contradictory results here are difficult to explain. It is not clear, for example, why stronger effects should be found with standardized achievement tests as criteria than with criterion-referenced measures. In any case, as Greenwood et al. (1979) have noted, effects should be shown for both type of criterion measure. It is also difficult to form any conclusions about whether some behavioral dimensions are more closely related to achievement than others, but it is worth noting in this connection that the positive effects obtained where composite mesures of academic survival skills formed from specific skills were employed. This is the same type of measure for which evidence of criterion-related validity was obtained with the correlational studies. In addition to these variations in type of criterion measure and type of behavior measure, there was variability among these studies with respect to subject characteristics, contextual variables, and design considerations. Some of these factors may play a role in this behavior-achievement relation, but it remains for future research to more fully explore that role.

Summary of the Review

The paper has focused on the behavioral measures used in recent classroom intervention research. A survey of the types of measures used in that research revealed a number of different observation systems. These systems differed in terms of the range of behaviors included and in terms of the specificity of response categories. The survey also revealed some variability in the precision with which the observation categories were operationalized and some inconsistencies in the way in which similarly labeled categories were defined from one system to another.

The next section of the paper presented an evaluation of these behavioral measures, an evaluation based on available empirical data. Several types of analyses were involved in this research, and all were shown to relate to certain key assumptions which have been made about the validity of these behavioral measures. The outcomes of the analyses yielded rather mixed results so far as the assumptions were concerned. There were cases where the data clearly supported both the criterion-related and construct validity of the measures. On the other hand, there were many failures to establish relations. There are two points to be kept in mind in considering these conclusions. First, the evaluation was based on a relatively small sample of studies; the issue of the validity of these measures has not been the object of much direct attention as yet. Second, some of the studies reviewed exhibited methodological or conceptual flaws which may have affected the adequacy of the validation tests. The sum of these points is that there are, as yet, no bases for any conclusive or final statements regarding the meaningfulness and relevance of these behavioral measures.

Recommendations

This review was prepared as a guide for those who make use of these observational measures in research and applied contexts and for those interested in research on the measures themselves. Two sets of recommendations will, therefore, be stated.

Implications for Use of the Measures

The first recommendation is that, where possible, existing observation schedules should be employed. There seems to be a tendency for researchers and practitioners to think their situation is unique, and that it is necessary for them to develop their own behavioral measures. As this survey has revealed, however, there is a wide range of observation schedules available and one or another of those schedules should be appropriate for most situations. This recommendation, if followed, should save time and effort on the part of the researcher. The practice might also contribute to the development of truly standardized behavioral measures where researchers take care in the collection and reporting of data (cf. Hartmann et al., 1979; Nelson & Bowles, 1975; Wasik & Loven, 1980).

A second recommendation is that care should be taken in developing precise operational definitions for the observational categories. The adequacy of these definitions has a direct impact on the quality of information collected in a study or project. Perhaps more important, however, is the fact that the precision of definitions has an impact on the ease with which others may interpret, evaluate, and replicate a study.

Third, researchers and practitioners should attend more closely to the measurement properties of their observation instruments than has been the

case in the past. There are a number of problem areas here. Thus, Wasik and Loven (1980) have shown that a number of problems exist with respect to the assessment of the reliability of these classroom measures through interobserver agreement procedures. Also worthy of note in this connection is Cone and Foster's (1982) discussion of other aspects of reliability assessment, including generalizability over time and over settings, which tend to be neglected in the use of these measures. Finally, as this review has sought to document, there are a number of questions which remain open with respect to the validity of these classroom observation schedules. These various questions and limitations must be acknowledged where making use of these measures as assessment devices in applied settings and where drawing conclusions from research based on the measures.

Implications for Future Research on the Measures

The first recommendation is that more research is needed in relating these observational measures of classroom behaviors to alternative types of measures. It is not sufficient to depend on content validity as we have in the past; rather, it is essential to establish the meaning of these measures through empirical procedures. The exercise is probably more critical in the case of the global and composite types of measures than it is for the specific measures, but all types should be subjected to empirical scrutiny. The recommended strategy in this case involves relating parallel measures of categories or dimensions through multitrait-multimethod designs (Cone, 1979; Cone & Foster, 1982; Gresham, 1982). It may also be advised that, while teacher judgment measures should continue to be used, other types of alternative measures should be explored, including peer and self ratings.

Further correlational research in which behavioral measures of classroom behavior are related to achievement indices constitutes the second recommendation. Three specific issues should be addressed in this research. First, there should be further efforts to assess the relative predictability of various specific behaviors. This would represent a continuation of the search for critical academic survival skills begun a number of years ago by Cobb (1970). Second, there is a clear need for further empirical investigations of the formation of composite categories from specific categories (Cone, 1981; Foster & Cone, 1980; Haynes, 1979). Efforts to form composites through multiple regression procedures have met with some success and should be continued. Third, more efforts should be made within this correlational approach to identify moderator variables. It seems clear that there is no single set of academic survival skills; rather, the behavior-achievement relation must vary across situations and persons. It is important to identify these critical variables.

The final recommendation is that more efforts should be made to explore the behavior-achievement relation within experimental designs. Past efforts along these lines have met with only limited success. However, more research is needed and two possible directions for this research will be indicated here. First, there is a need for experimental designs incorporating both standardized achievement tests and criterion-referenced tests as dependent measures, with the measures collected over a relatively long period of time. A similar kind of suggestion has been made by Greenwood et al. (1979). Second, researchers are advised to explore more closely the behavior-achievement links within these experiments. This would likely involve the use of one of the multivariate designs.

Considerable progress has been made over the past 15 years or so in the development of behavioral intervention strategies and in the investigation of the dynamics of classroom processes. It seems safe to assume, however, that future progress in these areas will be paced to a large extent by improvements in our measuring instruments. This paper constitutes a plea for more attention to one class of these behavioral measures, those focusing on the classroom behavior of the pupil.

References

Abikoff, H., Gittelman-Klein, R., & Klein, D. F. (1977). Validation of a classroom observation code for hyperactive children. Journal of Consulting and Clinical Psychology, 45, 772-783.

Abikoff, H., Gittelman, R., & Klein, D. F. (1980). Classroom observation code for hyperactive children: A replication of validity. Journal of Consulting and Clinical Psychology, 48, 555-565.

Ayllon, T., Layman, D., & Kandel, H. J. (1975). A behavioral-educational alternative to drug control of hyperactive children. Journal of Applied Behavior Analysis, 8, 137-146.

Ayllon, T., & Roberts, M. D. (1974). Eliminating discipline problems by strengthening academic performance. Journal of Applied Behavior Analysis, 7, 71-76.

Blunden, D., Spring, C., & Greenberg, L. M. (1974). Validation of the classroom behavior inventory. Journal of Consulting and Clinical Psychology, 42, 84-88.

Bolstad, O. D., & Johnson, S. M. (1977). The relationship between teachers' assessment of students and the students' actual behavior in the classroom. Child Development, 48, 570-578.

Boyd, L. A., Keilbaugh, W., & Axelrod, S. (1981). The direct and indirect effects of positive reinforcement on on-task behavior. Behavior Therapy, 12, 80-92.

Broughton, S. F., & Lahey, B. B. (1978). Direct and collateral effects of positive reinforcement, response cost, and mixed contingencies for academic performance. Journal of School Psychology, 16, 126-136.

Cairns, R. B., Green, J. A. (1979). How to assess personality and social patterns: Observations or ratings? In R. B. Cairns (Ed.), The analysis of social interactions (pp. 209-226). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cameron, M. I., & Robinson, V. (1980). Effects of cognitive training on academic and on-task behavior of hyperactive children. Journal of Abnormal Child Psychology, 8, 405-419.

Center, D. B., Deitz, S. M., & Kaufman, M. E. (1982). Student ability, task difficulty, and inappropriate classroom behavior. Behavior Modification, 6, 355-373.

Cobb, J. A. (1969). The relationship of observable classroom behaviors to achievement of fourth grade pupils. Unpublished doctoral dissertation, University of Oregon.

Cobb, J. A. (1970). Survival skills and first grade academic achievement. Report No. 1, Center at Oregon for Research in the Behavioral Education of the Handicapped. University of Oregon, Eugene, Oregon, 1970.

Cobb, J. A. (1972). Relationship of discrete classroom behaviors to fourth-grade academic achievement. Journal of Educational Psychology, 63, 74-80.

Cobb, J. A., & Hops, H. (1973). Effects of academic survival skill training on low achieving first graders. The Journal of Educational Research, 67, 108-113.

Cone J. D. (1979). Confounded comparisons in triple response mode assessment research. Behavioral Assessment, 1, 85-95.

Cone, J. D. (1981). Psychometric considerations. In M. Hersen & A. S. Bellack (Eds.), Behavioral assessment: A practical handbook (pp. 38-68). New York: Pergamon Press.

Cone, J. D. (1982). Validity of direct observation assessment procedures. In D. P. Hartmann (Ed.), Using observers to study behavior: New directions for methodology of social and behavioral science (pp. 67-79). San Franciso: Jossey-Bass.

Cone, J. D., & Foster, S. L. (1982). Direct observation in clinical psychology. In P. Kendall & J. Butcher (Eds.), Handbook of research methods in clinical psychology (pp. 311-354). New York: Wiley.

Darch, L. B., & Thorpe, H. W. (1977). The principal game: A group consequence procedure to increase classroom on-task behavior. Psychology in the Schools, 14, 341-347.

Deitz, S. M., Slack, D., Schwarzumeller, E. B., Wilander, A. P., Weatherly, T., & Hillard, G. (1978). Reducing inappropriate behavior in special classrooms by reinforcing average interresponse times: Intervall DRL. Behavior Therapy, 9, 37-46.

Dunkin, M. J., & Biddle, B. J. (1974). The study of teaching. New York: Holt, Rinehart & Winston.

Eastman, B. G., & Rasbury, W. C. (1981). Cognitive self-instruction for the control of impulsive classroom behavior: Ensuring the treatment package. Journal of Abnormal Child Psychology, 9, 381-387.

Emery, R. E., & Marholin, D. (1977). An applied behavior analysis of delinquency: The irrelevancy of relevant behavior. American Psychologist, 32, 860-873.

Ferritor, D. E., Buckholdt, D., Hamblin, R. L., & Smith, L. (1972). The non-effects of contingent reinforcement for attending behavior on work accomplished. Journal of Applied Behavior Analysis, 5, 7-17.

Foster, S. L., & Cone, J. D. (1980). Current issues in direct observation. Behavioral Assessment, 2, 313-338.

Friedling, C., & O'Leary, J. G. (1979). Effects of self-instructional training on second- and third-grade hyperactive children: A failure to replicate. Journal of Applied Behavior Analysis, 12, 211-219.

Green, [illegible]D., Beck, S. J., Forehand, R., & Vosk, B. (1980). Validity of teacher nominations of child behavior problems. Journal of Abnormal Child Psychology, 8, 397-404.

Greenwood, C. R., Hops, H., & Walker, H. M. (1977a). The program for academic survival skills (PASS): Effects on student behavior and achievement. Journal of School Psychology, 15, 25-35.

Greenwood, C. R., Hops, H., & Walker, H. M. (1977b). The durability of student behavior change: A comparative analysis at follow-up. Behavior Therapy, 8, 631-638.

Greenwood, C. R., Hops, H., Walker, H. M., Guild, J. J., Stokes, J., Young, K. R., Keleman, K. S., & Willardson, M. (1979). Standardized classroom management program: Social validation and replication studies in Utah and Oregon. Journal of Applied Behavior Analysis, 12, 235-253.

Gresham, F. M. (1982). A model for the behavioral assessment of behavior disorders in children: Measurement considerations and practical application. Journal of School Psychology, 20, 131-144.

Hall, V. C., Huppert, J. W., & Levi, A. (1977). Attention and achievement exhibited by middle- and lower-class black and white elementary school boys. Journal of Educational Psychology, 69, 115-120.

Hallahan, D. P., Lloyd, J. W., Kneedler, R. D., & Marshall, K. J. (1982). A comparison of the effects of self- versus teacher-assessment of on-task behavior. Behavior Therapy, 13, 715-723.

Hartmann, D. P., Roper, B. L., & Bradford, D. C. (1979). Some relationships between behavioral and traditional assessment. Journal of Behavioral Assessment, 1, 3-21.

Hartup, W., Glazer, J., & Charlesworth, R. (1967). Peer reinforcement and sociometric status. Child Development, 38, 1017-1024.

Hay, W. M., Hay, L. R., & Nelson, R. O. (1977). Direct and collateral changes in on-task and academic behavior resulting from on-task versus academic contingencies. Behavior Therapy, 8, 431-441.

Haynes, S. N. (1979). Behavioral variance, individual differences, and trait theory in a behavioral construct system: A reappraisal. Behavioral Assessment, 1, 41-49.

Haynes, S. N., & Kerns, R. D. (1979). Validation of a behavioral observation system. Journal of Consulting and Clinical Psychology, 47, 397-400.

Herbert, J., & Attridge, C. (1975). A guide for developers and users of observation systems and manuals. American Educational Research Journal, 12, 1-20.

Hoge, R. D. (in press). Psychometric properties of teacher judgment measures of pupil aptitudes, classroom behaviors, and achievement levels. Journal of Special Education.

Hoge, R. D., & Luce, S. (1979). Predicting academic achievement from classroom behavior. Review of Educational Research, 49, 479-496.

Hops, H., & Cobb, J. A. (1974). Initial investigations into academic survival skill training, direction instruction, and first-grade achievement. Journal of Educational Psychology, 66, 548-553.

Hops, H., Walker, H. M., Hernandez Fleischman, D., Nagoshi, J. T., Omura, R.T., Skindrud, K., & Taylor, J. (1978). Class: A standardized in-class program for acting-out children: II. Field test evaluation. Journal of Educational Psychology, 70, 636-644.

Hudgins, B. B. (1967). Attending and thinking in the classroom. Psychology in the Schools, 4, 211-216.

Hundert, J., Bucher, B., & Henderson, M. (1976). Increasing appropriate classroom behavior and academic performance by reinforcing correct work alone. Psychology in the Schools, 13, 195-200.

Jones, F. H., Fremouw, W., & Carples, S. (1977). Pyramid training of elementary school teachers to use a classroom management "skill package". Journal of Applied Behavior Analysis, 10, 239-253.

Karweit, N., & Slavin, R. E. (1981). Measurement and modeling choices in studies of time and learning. American Educational Research Journal, 18, 157-171.

Karweit, N., & Slavin, R. E. (1982). Time-on-task: Issues of timing, sampling, and definition. Journal of Educational Psychology, 74, 844-851.

Kirby, F. D., & Shields, F. (1972). Modification of arithmetic response rate and attending behavior in a seventh-grade student. Journal of Applied Behavior Analysis, 5, 79-84.

Klein, R. D. (1979). Modifying academic performance in the grade school classroom. Progress in Behavior Modification, 8, 293-321.

Lahaderne, H. M. (1968). Attitudinal and intellectual correlates of attention: A study of four sixth-grade classrooms. Journal of Educational Psychology, 59, 320-324.

Lahey, B. B., Green, K. D., & Forehand, R. (1980). On the independence of ratings of hyperactivity, conduct problems, and attention deficits in children: A multiple regression analysis. Journal of Consulting and Clinical Psychology, 48, 566-574.

Lipe, D., & Jung, J. M. (1971). Manipulating incentives to enhance school learning. Review of Educational Research, 41, 249-280.

Lobitz, W. C., & Burns, W. J. (1977). The "least intrusive intervention" strategy for behavior change procedures: The use of public and private feedback in school classrooms. Psychology in the Schools, 14, 89-94.

Loney, J., Weissenburger, F. E., Woolson, R. F., & Lichty, E. (1979). Comparing psychological and pharmacological treatments for hyperactive boys and their classmates. Journal of Abnormal Child Psychology, 7, 133-143.

Luce, S. R., & Hoge, R. D. (1978). Relations among teacher rankings, pupil-teacher interactions, and academic achievement: A test of the teacher expectancy hypothesis. American Educational Research Journal, 15, 489-500.

Main G. C., & Munro, B. C. (1977). A token reinforcement program in a public junior-high school. Journal of Applied Behavior Analysis, 10, 93-94.

Marholin, D., & Steinman, W. M. (1977). Stimulus control in the classroom as a function of the behavior reinforced. Journal of Applied Behavior Analysis, 10, 465-478.

Marlowe, R. H., Madsen, C. H., Bowen, C. E., Reardon, R. C., & Logue, P. E. (1978). Severe classroom behavior problems: Teachers or counsellors. Journal of Applied Behavior Analysis, 11, 53-66.

McKinney, J. D., Mason, J., Perkerson, K., & Clifford, M. (1975). Relationship between classroom behavior and academic achievement. Journal of Educational Psychology, 67, 198-203.

Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. Psychological Bulletin, 89, 575-588.

Nelson, C. M. (1971). Techniques for screening conduct disturbed children. Exceptional Children, 37, 501-507.

Nelson, R. O. (1983). Behavioral assessment: Past, present, and future. Behavioral Assessment, 5, 195-206.

Nelson, R. O., & Bowles, P. E. (1975). The best of two worlds — Observations with norms. Journal of School Psychology, 13, 3-9.

Nelson, R. O., & Hayes, S. C. (1979). Some current dimensions of behavioral assessment. Behavior Assessment, 1, 1-16.

O'Leary, K. D. (1972). Behavior modification in the classroom: A rejoinder to Winett and Winkler. Journal of Applied Behavior Analysis, 5, 505-511.

Page, D. P., & Edwards, R. P. (1978). Behavior change strategies for reducing disruptive classroom behavior. Psychology in the Schools, 15, 413-418.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R. Travers (Ed.), Second handbook of research on teaching (pp. 122-183). Chicago: Rand McNally.

Samuels, S. J., & Turnure, J. E. (1974). Attention and reading achievement in first-grade boys and girls. Journal of Educational Psychology, 66, 29-32.

Sherman, J. A., & Bushell, D. (1975). Behavior modification as an educational technique. F. D. Horowitz (Ed.), Review of child development research, volume 4 (pp. 409-462). Chicago: University of Chicago Press.

Soli, S. D., & Devine, V. T. (1976). Behavioral correlates of achievements: A look at high and low achievers. Journal of Educational Psychology, 68, 335-341.

Waksman, S. A. (1979). An evaluation of social learning procedures designed to aid students with conduct problems. Psychology in the Schools, 16, 416-421.

Walker, H. M., & Hops, H. (1976). Increasing academic achievement by reinforcing direct academic performance and/or facilitative nonacademic responses. Journal of Educational Psychology, 68, 218-225.

Warner, S. P., Miller, F. D., & Cohen, M. W. (1977). Relative effectiveness of teacher attention and the "good behavior game" in modifying disruptive classroom behavior. Journal of Applied Behavior Analysis, 10, 737.

Wasik, B. H., & Loven, M. D. (1980). Classroom observational data: Sources of inaccuracy and proposed solutions. Behavioral Assessment, 2, 211-227.

Werry, J. S., & Quay, H. C. (1969). Observing the classroom behavior of elementary school children. Exceptional Children, 35, 461-469.

Whalen, C. K., Henker, B., Collins, B. E., Finck, D., & Dotemoto, S. (1979). A social ecology of hyperactive boys: Medication effects in structured classroom environments. Journal of Applied Behavior Analysis, 12, 65-81.

Winett, R. A., & Roach, D. (1973). The effects of reinforcing academic performance on social behavior: A brief report. Psychological Record, 23, 391-396.

Winett, R. A., & Winkler, R. C. (1972). Current behavior modification in the classroom: Be still, be quiet, be docile. Journal of Applied Behavior Analysis, 5, 499-504.

Witt, J. C., & Adams, R. M. (1980). Direct and observed reinforcement in the classroom. Behavior Modification, 4, 321-336.

Wolraich, M., Drummond, T., Salomon, M. K., O'Brien, M. L., & Sivage, C. (1978). Effects of methylphenidate alone and in combination with behavior modification procedures on the behavior and academic performance of hyperactive children. Journal of Abnormal Child Psychology, 6, 149-161.

Zentall, S. S. (1980). Behavioral comparisons of hyperactive and normally active children in natural settings. Journal of Abnormal Child Psychology, 8, 93-109.

Author Notes

The preparation of this paper was supported in part by a grant from the Pickering Foundation. Gratitude is expressed to D. Andrews, B. Laver, and P. D. McCormack for their comments on an earlier draft of this paper.

Footnotes

[1]The following journals were included in the survey: American Educational Research Journal (vol. 14-20), Behavior Modification (vol. 1-7), Behavior Research and Therapy (vol. 15-21), Behavior Therapy (vol. 8-14), Behavioral Assessment (vol. 1-5), Journal of Abnormal Child Psychology (vol. 5-11), Journal of Applied Behavior Analysis (vol. 10-16), Journal of Consulting and Clinical Psychology (vol. 45-51), Journal of Educational Psychology (vol. 69-75), Journal of School Psychology (vol. 15-21), Journal of Special Education (vol. 11-17), and Psychology in the Schools (vol. 14-20). Studies involving retarded or special clinical groups were not included in the survey.

[2]A related literature also exists in which measures of pupil time-on-task are related to achievement indices (e.g., Karweit & Slavin, 1981, 1982). That literature falls somewhat outside the scope of this review, but it is called to the reader's attention.

[3]Only experimental studies including both achievement indices and observational measures of classroom behavior were included in the review.

[4]Ferritor, Buckholt, Hamblin, and Smith (1972) found significant effects only for conditions in which behavioral and performance interventions were combined.

Table 1

Summary of Behavioral Measures

Category Name	Investigation
a. (i) "on-task"; (ii) "off task"	Boyd, Keilbaugh, & Axelrod (1981); Broughton & Lahey (1978); Cameron & Robinson (1980); Darch & Thorpe (1977); Eastman & Rasbury (1981); Friedling & O'Leary (1979); Hallahan, Lloyd, Kneedler, & Marshall (1982); Hay, Hay, & Nelson (1977); Lobitz & Burns (1977); Loney, Weissenburger, Woolson, & Lichty (1979); Marlowe, Madsen, Bowen, Reardon, & Logue (1978)
b. (i) "on-task"; (ii) "disruptive"; (iii) "neutral"	Marholen & Steinman (1977)
c. (i) "disruptive"	Deitz, Slack, Schwarzmueller, Wallender, Weatherly, & Hilliard (1978); Warner, Miller, & Cohen (1977)

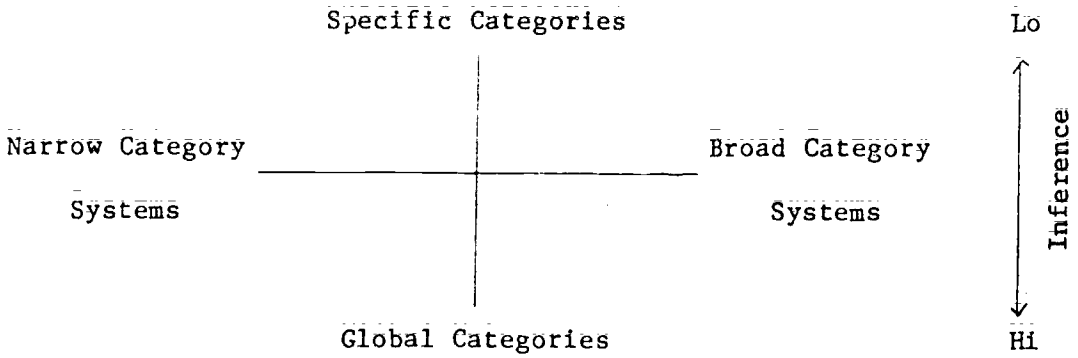
	Category Name	Investigation
d.	(i) "appropriate"; (ii) "inappropriate"	Center, Deitz, & Kaufman (1982); Waksman (1979); Witt & Adams (1980)
e.	(i) "appropriate"; (ii) "out-of-seat, inappropriate"; (iii) "talk-to-peer, inappropriate"; (iv) "talk-to-teacher, negative"; (v) "other off-task"	Page & Edwards (1978)
f.	(i) "talk-to-neighbor, inappropriate"; (ii) "out-of-seat, inappropriate"	Jones, Fremouw, & Carples (1977)
g.	(i) "task attention"; (ii) "out-of-chair"; (iii) "movement"; (iv) "fidget"; (v) "negative verbalization"; (viii) "translocation"; (ix) "noise"; (x) "physical contact, negative"; (xi) "physical contact, positive"; (xii) "social initiation"; (xiii) "high energy"; (xiv) "disruption"; (xv) "stand-out, negative"; (xvi) "sudden change"; (xvii) "grimace"; (xviii) "accident"; (xix) "ignore"; (xx) "bystand"	Whalen, Henker, Collins, Finck, & Dotemoto (1979)

	Category Name	Investigation
h.	(i) "attention"; (ii) "working"; (iii) "compliance"; (iv) "talk-to-peer, positive"; (v) "volunteering"; (vi) "self-stimulation"; (vii) "out-of-chair"; (viii) "look around"; (ix) "not attending"; (x) "play".	Greenwood, Hops, & Walker (1977a, 1977b)[a]; Greenwood, Hops, Walker, Guild, Stokes, Young, Keleman, & Willardson (1979)[a]
i.	(i) "attend"; (ii) "academic talk"; (iii) "work"; (iv) "volunteer"; (v) "management"; (vi) "approve"; (vii) "play"; (viii) "irrelevant talk"; (ix) "look around"; (x) "inappropriate locale"; (xi) "disruptive"; (xii) "physical, negative"; (xiii) "disapproval".	Hops, Walker, Fleischman, Nagoshi, Omura, Skindrud & Taylor (1978)
j.	(i) "motor behavior, inappropriate"; (ii) "aggression"; (iii) "disturbing property"; (iv) "disruptive noise"; (v) "turning around"; (vi) "verbalization, inappropriate"; (vii) "inappropriate task".	Main & Munro (1977)
k.	(i) "out-of-seat"; (ii) "inappropriate vocalization"; (iii) "nonattending"; (iv) "peer interaction"; (v) "fidgeting".	Wolraich, Drummond, Salomon, O'Brien, & Sivage (1978)

Notes: a. There were some variations among these three studies with respect to the number and labeling of categories.

Figure Caption

Figure 1. Dimensions for describing the category systems.


Specific Categories
Narrow Category
Systems
Broad Category
Systems
Global Categories
Lo
Inference
Hi